

WHAT IS CLAIMED IS:

- sub  
A17
1. A data compressing apparatus for generating code data from a character train stream constructed by a document including tags, comprising:
- 5 a tag information separating unit for separating the identified tag from said character train stream and outputting as tag information;
- a tag code replacing unit for arranging a tag code for identification to a position of the character train stream in which the tag was separated by said tag information separating unit; and
- 10 a character train coding unit for coding the character train stream including the tag code outputted from said tag code replacing unit and outputting a code stream.
- 15
- sub  
B17
2. An apparatus according to claim 1, wherein said tag code replacing unit arranges a predetermined fixed code as said tag code to the position of the character train stream in which the tag was separated.
- 20
3. An apparatus according to claim 1, wherein said tag code replacing unit arranges a tag code indicative of an appearing order of the tag separated by said tag information separating unit to the position of the character train stream in which the tag was separated.
- 25
- 0360355 071399  
66E7/0" 5500000

4. An apparatus according to claim 1, further comprising:

a tag information storing unit for storing the tag information separated by said tag information separating unit;

a code storing unit for storing the code data formed by said character train coding unit; and

a code switching unit for selecting the tag information stored in said tag information storing unit and the code data stored in said code storing unit and outputting the selected tag information or code data.

5. An apparatus according to claim 1, wherein said character train coding unit comprises:

a dictionary storing unit for storing a dictionary in which a character train serving as a processing unit when compressing has been registered; and

a character train comparing unit for comparing a partial character train in the character train stream from said tag code replacing unit with the registration character train in said dictionary storing unit, thereby detecting a partial character train which coincides with said registration character train, allocating a predetermined code every said detected partial character train, and outputting a resultant character train.

0932255 07199 66ET/0" 55225560

Sub  
B17

6. An apparatus according to claim 1, further comprising a tag information compressing unit for compressing the tag information separated by said tag information separating unit.

5

7. An apparatus according to claim 1, further comprising:

10 a tag dictionary storing unit for storing a dictionary in which a tag character train in the tag information serving as a processing unit when compressing has been registered; and

15 a tag character train comparing unit for comparing the partial character train of the character train stream included in the tag information separated by said tag information separating unit with the registration character train in said tag dictionary storing unit, thereby detecting a partial character train which coincides with said registration character train, allocating a predetermined code every said  
20 detected partial character train, and outputting a resultant character train.

8. An apparatus according to claim 4, further comprising a tag position detecting unit for detecting  
25 a position of the tag in the code data formed by said character train coding unit,

and wherein both the tag information separated by

0935355.071399

said tag information separating unit and designation information of the tag position detected by said tag position detecting unit are stored in said tag information storing unit.

5

9. An apparatus according to claim 8, wherein said tag position detecting unit detects the code amount from the head of a document or a specific tag and stores it together with the tag information into said tag information storing unit.

10

50  
A2  
10. A data reconstructing apparatus for reconstructing character train data from a code stream including tag information separated from a character train stream of a document including tags and code data obtained by encoding a character train stream in which a tag code has been arranged at a position of the separated tag, comprising:

15

a tag information separating unit for separating said tag information and said code data from said code stream;

20

a tag information storing unit for storing the tag information separated by said tag information separating unit; and

25

a character train reconstructing unit for reconstructing the character train data including the character train and the tag code from said code data

and, thereafter, replacing said tag code by the tag information in said tag information storing unit.

11. An apparatus according to claim 10, wherein said  
5 character train reconstructing unit comprises:

a dictionary storing unit for storing a dictionary in which a reconstruction character train corresponding to a code of the character train serving as a processing unit when reconstructing has been  
10 registered;

a character train comparing unit for separating a code of the character train serving as a reconstruction unit from said code stream and reconstructing the original character train with reference to said  
15 dictionary storing unit; and

a character train replacing unit for replacing the tag code reconstructed by said character train comparing unit by the tag information in said tag information storing unit.  
20

12. An apparatus according to claim 10, further comprising a tag information reconstructing unit for reconstructing compression data of the tag information stored in said tag information storing unit.

25 13. An apparatus according to claim 10, further comprising:

00363355 074399  
668740 9522520

a tag dictionary storing unit for storing a dictionary in which a reconstruction character train corresponding to a code of a tag character train serving as a processing unit when reconstructing has  
5 been registered; and

a tag character train comparing unit for separating a code of the tag character train serving as a reconstruction unit from the tag information separated by said tag information separating unit and  
10 reconstructing the original tag character train with reference to said dictionary storing unit.

14. A data compressing method of generating code data from a character train stream constructed by a document  
15 including tags, comprising:

a tag information separating step of separating the identified tag from said character train stream and outputting as tag information;

a tag code replacing step of arranging a tag code  
20 for identification to a position of the character train stream in which the tag was separated in said tag information separating step; and

a character train coding step of coding the character train stream including the tag code outputted  
25 from said tag code replacing step and outputting a code stream.

0935355.071399  
66ET20 552560

Sub  
A37

add  
A37